# Ngee Ann Polytechnic Rebrand: Child's Play

# School of InfoComm Technology

**Data Exploration & Analysis Assignment**

Diploma in DS

April 2024 Semester

**ASSIGNMENT 2**

(40% of DEA Module)

8th July 2024 – 11th August 2024

**Submission Deadline:**

**Presentation: Week 17 during class**

**Report and files: 11th August 2024 (Sunday), 11:59PM**

|  |  |  |
| --- | --- | --- |
| **Tutorial Group** | **:** | **T01** |
| **Student Name** | **:** | Ang Wei Xuan Senderos |
| **Student Number** | **:** | S10257398 |

**Penalty for late submission:**

Marks will be deducted every calendar day after the deadline.

**NO** submission will be accepted after 18th August 2024 (Sunday), 11:59PM.

Contents

[**Abstract / Overview** 3](#_Toc174207516)

[**Creating dashboard to answer business question (3.1)** 4](#_Toc174207517)

[**Further Analysis (3.2)** 21](#_Toc174207518)

[**Time Series Forecasting Model** 31](#_Toc174207519)

[**Data Mining** 34](#_Toc174207520)

[**Summary** 37](#_Toc174207521)

[**References** 39](#_Toc174207522)

# **Abstract / Overview**

This report delves into the creation and analysis of four critical dashboards designed to answer key business questions posed by the CEO of Flipkart, as well as an additional self-posed business question that provides further strategic insights. Each dashboard is meticulously crafted to address specific areas of business performance, ranging from sales performance to customer behaviour across different regions in India. The report highlights the methodologies used, including data cleaning, relationship establishment between tables, and the creation of custom measures in Power BI.

Further analysis is conducted through univariate, bivariate, and multivariate techniques, alongside data mining methods like K-means clustering and logistic regression. This deeper analysis focuses on identifying the factors that most significantly impact profitability, providing actionable insights that go beyond surface-level observations. The findings are summarized to highlight key trends and patterns, while recommendations are made to help Flipkart strengthen its competitive position in the e-commerce industry.

This report not only captures the technical execution of data analysis but also emphasizes the importance of data-driven decision-making in real-world business scenarios, equipping Flipkart with the strategic tools needed to navigate the complexities of the e-commerce landscape.

# **Creating dashboard to answer business question (3.1)**

**Additional Tables Created**

While beginning to create the dashboards, a new table named ‘Date’ was created with the following code:

A computer code with numbers and symbols

Description automatically generated

Despite having a date hierarchy already existing in the ‘List of Order Details’ table, creating a custom date table in Power BI allowed me to have more granular control, customizing how months, are represented such as Jan instead of the full abbreviation. This table has new features such as Day of Week, providing additional dimensions for analysis. Furthermore, using a single, consistent date table helps to avoid discrepancies that might arise from different tables having varying date formats.

Another table called ‘CategoryLookup’ which contains the 3 distinct product categories was created.

**Data Cleaning**

Before establishing relationships between each table, removal of null values was done in the ‘List of Order Details’.

**Establishing Relationship between the different tables**

A screenshot of a computer

Description automatically generated

List of Orders – Order Details (1-to-Many) Relationship on ‘Order ID’

List of Orders (‘Order Date’) – Date (‘Date’) (Many-to-One) Relationship

List of Orders (‘Order Date’) – Sales Target (‘Month of Order Date’) (Many-to-Many) Bidirectional Relationship

Order Details (‘Category’) – CategoryLookup (‘Category’) (Many-to-One) Relationship

Sales target (‘Category’) – CategoryLookup (‘Category') (Many-to-One) Relationship

**Q1 Dashboard Overview & Creation Process**

A screenshot of a computer screen

Description automatically generated

A close up of numbers

Description automatically generated

This KPI shows the month with most sales during the period of Q2/2018 to Q1/2019.

Formula:

A screenshot of a computer code

Description automatically generated

A number with red numbers

Description automatically generated

This KPI, 'Sales Target Achieved (%),' measures the percentage of the company's sales target reached within a specific period. It was calculated using the following formula:



This KPI is crucial for assessing whether the company meets its sales goals. In this specific visual, Flipkart achieved approximately 99% of its sales target for the period from Q2/2018 to Q1/2019. Tracking this metric helps the company make informed decisions regarding performance evaluation and strategic planning.

A white background with black and white clouds

Description automatically generated

These two KPIs display the target sales amount for a specific period, whether it's for a single month or across several months, alongside the actual sales amount. The target measure is derived from the 'Sales Target' table, while the actual sales amount uses the 'Sales' measure. Together, they provide a quick comparison and insight into performance against targets.

A graph with a line and numbers

Description automatically generated

Formula:

A white background with black text

Description automatically generatedA screenshot of a computer

Description automatically generated

This area chart shows the comparison of sales for each month with the previous month. For instance, in June 2018, its sales dipped by 17.12% from May. This chart provides context beyond simple KPIs by offering insights into trends in sales performance. By displaying both sales figures and growth rates together, it offers a more nuanced view of Flipkart’s performance.

A graph of a bar chart

Description automatically generated with medium confidence

This bullet chart provides a clear visual comparison of actual sales against the targets set by Flipkart for each month. By highlighting the months when sales met or fell short of the set goals, it allows for a quick performance assessment. Additionally, the use of blue and yellow colors makes this dashboard more accessible to individuals with color blindness, particularly those with Protanopia and Deuteranopia.

A graph with a line

Description automatically generated

This line chart visualizing Order Quantity over time provides essential insights for optimizing inventory management and labor costs. By monitoring these trends, Flipkart can anticipate demand fluctuations, adjust inventory levels, and avoid overstocking or stockouts. It also helps identify periods of low activity, enabling adjustments in staffing levels to reduce labor costs.

**Q2 Dashboard Overview & Creation Process**

A screenshot of a computer

Description automatically generated

A black text and numbers

Description automatically generated with medium confidence

Formula:

A white background with black text

Description automatically generated

The 'Average Profit Per Unit of Product Sold' KPI specifically helps pinpoint the profitability of each product, revealing whether certain items contribute disproportionately to overall profits or losses. Unlike broader KPIs that might aggregate profits, this metric allows Flipkart to identify underperforming products and adjust pricing strategies. For example, if profit for each electronic product is higher than furniture products, Flipkart can leverage the opportunity to sell more electronic products to gain greater profits.

A number with text on it

Description automatically generated

Formula:

A white background with black text

Description automatically generated

This KPI card displays the profit margin, which can be filtered by category or viewed for all categories. It's crucial for determining whether it aligns with the e-commerce industry’s standard profit margins. If certain periods are below the average by 10%, Flipkart can investigate potential issues, such as pricing strategy, cost management, or promotional effectiveness, and take corrective actions. When filtered by category, Flipkart can identify specific product lines with potential for profit margin improvement and focus efforts on optimizing pricing, reducing costs, or enhancing marketing strategies to boost profitability in those areas.

A blue and yellow bar chart

Description automatically generated

This bullet chart compares sales revenue to the set target for a given period, such as a month. In this visual, when filtered for January, it is evident that the clothing category did not meet its sales target. This helps identify underperforming categories or periods, crucial for understanding why overall sales revenue might not meet desired targets.

A screenshot of a computer screen

Description automatically generated

This tree map effectively addresses Q2: ‘What are the top 5 products?’. This insight allows for more effective targeting of marketing activities, ensuring promotional efforts focus on the most popular items ordered. For example, Saree has a high quantity ordered. So, the inventory manager of Flipkart can take note of this and order a greater supply and reduce the supply of less-demanded items, reducing dead stock.

A graph with a curved line

Description automatically generated with medium confidence

Formula:

A computer screen shot of a code

Description automatically generated

A screenshot of a computer program

Description automatically generated

A white background with black text

Description automatically generated



A Pareto analysis chart is included to identify the top products contributing to sales, following the 80/20 rule. This chart helps prioritize marketing efforts on high-impact products, optimize inventory management, and enhance sales strategies by focusing on the most significant contributors to revenue. It ensures resources are directed towards the products that drive most sales, increasing overall business efficiency and effectiveness

**Q3 Dashboard Overview & Creation Process**

**A screenshot of a computer

Description automatically generated**

**A number with text on it

Description automatically generated**

Formula:

A computer screen shot of a program

Description automatically generated

This KPI card reveals the percentage of profits contributed by the top 5 states out of the total 19, highlighting the concentration of profitability in key regions. This insight allows Flipkart to focus on optimizing existing operations, enhancing customer experience, and increasing market share in already profitable regions. Additionally, it helps identify whether the Pareto principle, or 80/20 rule, applies, which can inform whether targeted strategies in these key states could disproportionately increase overall profitability.

A blue rectangles with white text

Description automatically generated

A new column called region was used.

Formula:

A screenshot of a computer code

Description automatically generated

I assigned union territories to specific regions in this visualization to provide a clearer and more comprehensive overview of profit distribution across broader geographical areas. This approach allows for a more accurate representation of regional performance, ensuring that no area is overlooked in the analysis.

This visual also had a drill down function into state and city using ‘Location Hierarchy’.

A screenshot of a computer

Description automatically generated

This visual offers stakeholders a detailed, layered understanding of profitability across various geographic levels, from broad regions to specific cities. The drill-down capability allows for precise identification of areas that are underperforming or excelling. For instance, if profits in the South region are lower, Flipkart can assess which specific cities or branches are contributing to the decline. This allows the company to strategically adjust stock levels, reducing inventory in poorly performing areas and reallocating resources to more profitable regions like the West, ultimately enhancing overall efficiency.

A screenshot of a computer

Description automatically generated

These bar charts reveal the top and bottom 5 locations by profit, directly answering the question in Q3: ‘**What are the top and bottom 5 locations in terms of profit?**’. This enables Flipkart's CEO to quickly identify disparities in profitability across states and cities. For instance, identifying locations with consistent net losses allows the company to analyze whether these areas suffer from low market demand, ineffective marketing strategies, or operational inefficiencies. Based on these insights, Flipkart can either implement targeted strategies to turn around these locations or consider reallocating resources to more profitable regions.

A blue screen with white text

Description automatically generated

This tree map visualizes the top 5 areas by items ordered, providing valuable insights into the most active regions and cities. By identifying these high-demand areas, Flipkart can strategically allocate inventory to boost sales. The ability to drill down into specific cities offers even more granular insights, enabling precise targeting and resource allocation. For example, if Madhya Pradesh has more items ordered there, inventory could be increased in that area to meet customers’ high demands.

**Q4 Dashboard Overview & Creation Process**

**Business Question:** How do Flipkart’s customers vary across different regions in India?

**Rationale:** Understanding regional variations in customer demographics and purchasing behaviors is critical for Flipkart's growth and strategic planning. Additionally, analyzing data on purchase frequency, cross-selling among products can reveal opportunities for targeted promotions and customer retention strategies.

**A screenshot of a computer

Description automatically generated**

A number of black numbers

Description automatically generated with medium confidence

Formula:











Having a KPI for Customer Lifetime Value helps Flipkart gauge the long-term profitability of each customer. It enables targeted marketing, better resource allocation, and informed decision-making to enhance overall business growth and customer retention.

A number with black text

Description automatically generated

Formula:

A computer code with text

Description automatically generated



This code calculates the number of repeated customers by counting unique customers who placed more than one order, ensuring the count isn't blank. Additionally, it calculates the ratio of repeat customers to those who made a single purchase. This metric is crucial for assessing customer loyalty, as a low ratio indicates a need for improved retention strategies. By identifying a low rate of repeat purchases, Flipkart can implement and promote loyalty programs, recognizing that relying solely on one-time customers is unsustainable for long-term growth.

A black numbers and a dollar sign

Description automatically generated

Formula:



This "Average Purchase Value" KPI is crucial for assessing the revenue generated per transaction, providing insights into customer spending habits and the overall value each customer brings to the business. This metric helps Flipkart identify trends in consumer behavior, evaluate the effectiveness of pricing strategies, and understand the impact of promotions or discounts. By tracking average purchase value, the company can focus on strategies to increase transaction sizes, such as bundling products or offering upsells.

A number on a white background

Description automatically generated A close up of numbers

Description automatically generated

Formula for Number of Orders: 

Formula for Unique Customers:



These KPIs offer a quick snapshot of order volume and customer diversity, crucial for understanding market reach and customer engagement, aiding in decision-making and strategy refinement.

A blue squares on a white background

Description automatically generated

This bar chart effectively addressed the mainq uestion for Q4: ‘**What is the distribution of Flipkart’s customer base across different locations?**’. This can be drilled down into its state and cities, offering critical insights into geographic demand. By analyzing customer concentration, Flipkart can strategically prioritize regions for targeted marketing efforts and resource deployment. Additionally, understanding customer spread can inform decisions about where to strengthen local presence or explore new opportunities, ensuring that each region receives appropriate attention to maximize customer engagement.

A graph of a bar chart

Description automatically generated with medium confidence

Formula:

A screenshot of a computer program

Description automatically generated

The chart indicates that customers are more likely to purchase clothing and furniture products when they buy electronics. This pattern suggests that electronics purchases often lead to additional sales in these categories, as customers may view these items as complementary to their electronics purchase. For instance, a customer buying a new electronic device might also choose to buy related furniture or clothing to enhance their experience. By understanding this cross-selling behavior, Flipkart can create targeted promotions, such as bundling deals that offer discounts on furniture and clothing when an electronic product is purchased, thereby maximizing the cross-sell potential and increasing the overall value of each transaction.

A graph with blue lines

Description automatically generated

This line chart uses the previous formulas mentioned above. It shows both the average purchase value and purchase frequency. By examining these metrics together, the company can assess whether increased purchase frequency correlates with higher average purchase values, which could indicate that repeat customers tend to place larger orders. This insight is valuable for developing customer retention strategies, as it suggests that encouraging repeat purchases could lead to increased revenue per customer, as seen from July to November. Additionally, identifying periods where these metrics diverge can help pinpoint when and why customer engagement or spending behaviour changes.

# **Further Analysis (3.2)**

**Problem Statement**

Flipkart's CEO currently lacks a clear understanding of the factors influencing the company’s profitability. This gap in knowledge is critical as the company has only recently begun utilizing data analytics to maintain its competitive edge in the e-commerce market. Addressing this issue is vital because profit directly impacts Flipkart’s financial sustainability and growth. In this analysis, the target variable 'Profit' is examined through univariate, bivariate, multivariate, and data mining techniques to uncover the underlying factors and trends that drive profitability. Understanding these factors will enable more informed decision-making to enhance business performance.

**Univariate Analysis**

**Exploratory Questions**

1. What is the trend of profits?

2. What is the distribution of profits?

3. What if the distribution of profits across different product categories?

4. What is the distribution of profits across days of the week?

5. What are the top 5 products that generate the most profits for Flipkart?

A blue line graph with white text

Description automatically generated

1. What is the trend of profits?

This line chart shows the profit Flipkart has made from April 2018 to March 2019. It provides the insight that profits in Flipkart have not been doing well from April to May, with all these months resulting in net losses. However, profits picked up from October onwards and has been slowly increasing. The trend highlights critical periods where interventions may be necessary, such as addressing the causes of early-year losses or capitalizing on the upward trend starting in October to sustain and further increase profitability.

A graph with blue lines

Description automatically generated

2. What is the distribution of profits?

The distribution of profits across Order ID provides critical insights into the variability and concentration of profitability at the transaction level. By analyzing this distribution, Flipkart can identify which orders contribute most significantly to overall profitability and detect outliers, such as orders resulting in substantial losses. This analysis is essential for refining pricing strategies, improving product selection, and optimizing sales tactics. Notably, approximately 30% of total orders generate a profit of $0-49, while some orders result in significant losses or high profits. Flipkart could possibly utilize sales strategy to gain greater profits per transaction.

A blue and orange pie chart

Description automatically generated

3. What if the distribution of profits across different product categories?

Analyzing profit distribution by category is crucial for understanding which product lines are most profitable. This insight helps Flipkart prioritize investments and optimize product offerings. The pie chart shows that the Clothing category generates the highest profit (46.6%), followed closely by Electronics (43.81%), with Furniture contributing the least (9.59%). This suggests that while Clothing and Electronics are strong profit drivers for Flipkart, **Furniture may need strategic improvements or better marketing**.

A blue squares with white text

Description automatically generated

4. What is the distribution of profits across days of the week?

The chart indicates that Thursdays and Saturdays are the most profitable days for Flipkart, while **Mondays show a net loss**. Flipkart can focus on enhancing promotions and inventory levels on high-profit days while addressing the challenges of lower profitability on Mondays.

A blue bar graph with numbers

Description automatically generated

5. What are the top 5 products that generate the most profits for Flipkart?

This bar chart identifies the top 5 products generating the highest profits for Flipkart. Printers lead the list with a significant margin, followed by Bookcases and Accessories. These high-profit products are critical for driving revenue. Highlighting the most profitable products enables Flipkart to prioritize these items in inventory management, marketing efforts, and pricing strategies. This insight helps in focusing resources on products that yield the highest financial returns, ensuring sustained profitability and competitive advantage

**Bivariate Analysis**

**Exploratory Questions**

1. What is the distribution of profits across different product categories, and how does it vary by days of the week?

2. What is the distribution of profits by the top 5 states by product category?

3. What is the distribution of profits by the bottom 5 states by product category?

4. What is the distribution of profits by the top 10 cities by product category?

5. What is the distribution of profits by the bottom 10 cities by product category?

6. What is the correlation between order quantity and profits by product category?

7. Does a higher purchase value correlate with higher profits?

A blue and orange bars

Description automatically generated

1. What is the distribution of profits across different product categories, and how does it vary by days of the week?

This chart highlights that Monday consistently shows negative profits across all categories, indicating potential challenges or low sales on this day. Additionally, Furniture products struggle on Sunday, despite higher profitability on Saturdays. Addressing these issues could help Flipkart improve overall profitability by adjusting strategies for specific days and categories.

Understanding how profits vary by product category and day of the week allows Flipkart to optimize its inventory and marketing strategies. For example, boosting promotions for Furniture products on Saturdays could capitalize on its higher profitability, while addressing the low or negative profits on Monday could prevent potential losses.

A blue and orange squares

Description automatically generated

2. What is the distribution of profits by the top 5 states by product category?

This chart highlights the profit distribution across the top 5 states by product category. Maharashtra leads in profits, with significant contributions from all categories. Electronics products dominate profits in Madhya Pradesh and Uttar Pradesh, while Delhi and West Bengal show lower overall profits with Electronics products and Clothing products as the primary driver. It is seen that **Furniture products are not doing well at Uttar Pradesh**, as they accumulated net losses. Understanding these variations allows Flipkart to tailor marketing and inventory strategies by state, focusing on the most profitable product categories for these states to maximize revenue.

A blue and orange squares

Description automatically generated

3. What is the distribution of profits by the bottom 5 states by product category?

This chart highlights the profit distribution across the bottom 5 states by product category. Previously, when evaluating branch performance, Tamil Nadu was observed to perform significantly worse than Punjab in terms of profits. This chart provides deeper insights into the cause, revealing that Furniture products are primarily responsible for the poor performances in Tamil Nadu and Andhra Pradesh. Additionally, Electronics have not been popular in Punjab, resulting in net losses.

A graph of a bar chart

Description automatically generated with medium confidence

4. What is the distribution of profits by the top 10 cities by product category?

This chart reveals the distribution of profits across the top 10 cities by product category. Electronic and Clothing products dominate as the leading category for most of these states, with furniture products only dominating in Bhopal in terms of profits. Understanding the distribution of profits by city and product category helps Flipkart identify which cities are most profitable and which product categories drive that profitability.

A screen shot of a graph

Description automatically generated

5. What is the distribution of profits by the bottom 10 cities by product category?

This chart reveals the distribution of profits across the bottom 10 cities by product category.

**Electronic Products**: Not performing well in Chandigarh, Kohima, Patna, Jaipur

**Clothing Products**: Not performing well in Kashmir, Jaipur, Ahmedabad, Chennai

**Furniture Products**: Not performing well in Lucknow, Hyderabad, Ahmedabad and Chennai

States with >=2 Product Categories being negative profits:

Jaipur, Ahmedabad and Chennai.

Understanding the distribution of profits by product category across the bottom 10 cities enables Flipkart to identify specific geographic areas where targeted interventions are needed.

A screen shot of a graph

Description automatically generated

6. What is the correlation between order quantity and profits by product category?

This scatter plot illustrates that while lower quantities can result in either high profits or significant losses, higher quantities tend to cluster around lower profits or smaller losses. This suggests that larger orders may involve bulk discounts or lower-margin items, leading to reduced profitability. For instance, in the Furniture category, even as quantities increase, the profits do not necessarily follow suit, and losses can still occur. Understanding this inverse relationship between quantity and profit is critical for Flipkart to reassess its pricing and bulk order strategies. By recognizing that high-quantity orders often lead to reduced profitability, the company can refine its approach to maximize margins and avoid potential losses.

A screen shot of a graph

Description automatically generated

7. Does a higher purchase value correlate with higher profits?

This scatter plot illustrates the relationship between **Order Value** and **Profit**. The majority of orders cluster around lower order values (below $2,000) with profits close to zero, indicating that many transactions are making little profits. Notably, as the order value increases, there are instances of both significant profits and losses. Only when transactions above $4,000, there generally tend to be high net profits. This suggests variability in profitability for higher-value orders, where not all high-value transactions guarantee high profits.

**Multivariate Analysis**

**Exploratory Questions**

1. How do sales correlate with profits across different regions and product categories? Are there any regions or product categories where high sales do not translate into high profits?

A screenshot of a graph

Description automatically generated

This bar chart shows the distribution of both sales and profits, across different regions of India and Product categories. It is observed that furniture products are not performing well in the South region of India, while electronic products are struggling in the East region as we can see that they accumulated negative profits. It is also seen that despite very high sales for furniture and clothing products in the West region of India, the profits generated is comparable to that of the East and North region. In the East, furniture products achieved similar profit levels with significantly lower sales figures, and in the North, clothing products achieved comparable profit levels with half of the West’s sales figures.

This visual is essential for identifying regional and product-specific performance gaps. It highlights where sales may not be translating into proportional profits, signalling areas for strategic intervention.

# **Time Series Forecasting Model**

**Getting the Dataset**

**A screenshot of a computer

Description automatically generated**

Merged the ‘List of Orders’ with ‘Order Details’ in Jupyter Notebook, using the 'Order ID' as the key. Next, aggregated the data by 'Profit' and 'Order Date,' summing the total profits for each date. The resulting data was then sorted by the 'Order Date' after converting the date format to ensure proper chronological order. It is then sent to my file path.

A graph with blue and orange lines

Description automatically generated

The reason why I used AR (1) model is due to the following:

A graph with blue dots and numbers

Description automatically generated

(An ACF Plot done on Jupyter Notebook)

A graph with blue dots and numbers

Description automatically generated

(A PACF Plot done on Jupyter Notebook)

**Rationale**

The decision to use an AR (1) model for forecasting Flipkart's profits is based on the observed characteristics of the time series data. Specifically, the gradual decay in the Autocorrelation Function (ACF) plot suggests that the data follows an autoregressive process, which is typical in time series with a correlation structure where observations are correlated with their own past values as the lag increases. This pattern helps identify the autoregressive (AR) component in ARIMA modelling, where the persistence of correlations diminishes gradually over time.

Moreover, the Partial Autocorrelation Function (PACF) plot, generated in Jupyter Notebook, provides further evidence to support this decision. The PACF plot shows a significant spike at lag 1, after which the spikes do not significantly exceed the confidence interval. This indicates that only the first lag has a meaningful impact on the current value, making it the primary contributor to the autoregressive process. As a result, this justifies the use of an AR (1) model, where the current profit value is primarily dependent on the immediate previous profit value.

# **Data Mining**

**Undirected Data Mining: K-means Clustering**

**A chart with many colored dots

Description automatically generated**

The number of clusters (3) was decided by using the elbow method, using a function called optimise\_k\_meansA graph with a line

Description automatically generated

Formula:

A screenshot of a computer

Description automatically generated

A screenshot of a computer program

Description automatically generated

A computer screen shot of a program

Description automatically generated

This K-means clustering with 3 clusters shows that the data points are grouped mainly around the lower Amount (sales) and Profit values. The clusters highlight distinct patterns in profitability relative to sales amounts. Most transactions cluster around low sales and low profits, indicating a concentration of smaller, less profitable transactions. Some data points, particularly in the higher sales and profits range, form a separate cluster, indicating more profitable transactions with higher sales.

**Directed Data Mining: Logistic Regression**

A screenshot of a computer program

Description automatically generated

This code performs a logistic regression analysis to predict whether a transaction will result in positive or negative profits. The ProfitBinary variable is created to label profits as 1 (positive) or 0 (negative). Several categorical features such as Category, Sub-Category, State, and City are encoded into numerical values to be used as input features (X), along with Amount and Quantity. The data is split into training and testing sets, and a logistic regression model is trained. The model achieves a model error of approximately 34%, indicating moderate ability to correctly classify transactions as profitable or not. The coefficients suggest the importance of each feature, with Category\_encoded being particularly significant.

# **Summary**

**Summarization of Findings from Dashboard 3.1**

Number of orders are increasing. July’s performance was the worst, failing to reach even half the target. Electronics category exceeded Flipkart’s sales expectations, while clothing fell noticeably short. Electronic products yield the highest profit by product. Top 5 products by quantity ordered are furnishings, handkerchiefs, sarees, stoles, and T-shirts. A little more than half of Flipkart’s subcategories contribute to more than 80% of total sales. The South region performed the worst, this is due to the states of Andhra Pradesh and Tamil Nadu. Electronics has the highest cross-sell sales, followed by Clothing and Furniture. This suggests that customers are more likely to purchase additional electronics when buying related products, indicating a strong cross-sell potential in this category. Purchase Frequency is stagnant, with only noticeable months being August and January, suggesting purchasing activity from repeat customers. Repeated Customer to One-Time Customer ratio from Q2/2018 to Q1/2019 is 21%.

**Summarization of Findings from Further Analysis 3.2**

**Univariate Analysis**: Profits in Flipkart are generally increasing monthly after the period of April to September, where it was preciously making net losses. Monday’s result in net losses for Flipkart, Furniture products generate the least profit among the other categories, with only 9.59%, compared to Electronics and Clothing with 43.81% and 46.6%. Top 5 profit drivers for products are printers, bookcases, accessories, trousers and stole.

**Bi-Variate Analysis**: All products, regardless of category suffer negative profits on Mondays, with Furniture performing badly on Sunday and Electronics on Tuesdays. Distribution of profits by Category for Top 5 states show how furniture products are not doing well in Uttar Pradesh. Tamil Nadu and Andhra Pradesh’s cause of bad performance is largely from its furniture products, accumulating net losses in the thousands. Punjab and Bhar’s poor performance are due to its electronic products. Further investigation found Chennai, a city in Tamil Nadu, accumulated huge net losses in its furniture and clothing products, hence the state of Tamil Nadu’s poor performance is because of Chennai. Lastly, order values that are $4,000 tend to generate huge profits.

**Multivariate Analysis**: Furniture struggles in the South, while electronics face losses in the East. Despite high sales in the West, furniture and clothing profits are on par with the East and North regions, where significantly lower sales still generate similar profits. Notably, clothing in the North achieves comparable profit levels to the West with only half the sales, suggesting inefficiencies in the West’s profit generation despite high sales.

**Recommendations**

Based on the insights derived from 3.1 and 3.2, some recommendations that Flipkart can take to enhance its competitive position in the e-commerce industry are:

**Leverage Cross-Selling opportunities**: It was seen that Electronics show the highest cross-sell potential. Flipkart should enhance its recommendation algorithms and bundling offers to capitalize on this behavior. Offering discounts on related electronic accessories or promoting combo deals during peak purchase periods could further increase sales and profitability in this category. Additionally, considering the high profitability of electronics, focused marketing efforts during high-profit days like Thursdays and Saturdays can amplify results.

**Divest from Chennai**: Chennai has consistently underperformed, with minimal profits generated only from electronic products, which are insufficient to offset the significant losses in furniture and clothing. Given this persistent underperformance, Flipkart should consider divesting from Chennai, reallocating resources to more profitable regions like Maharashtra and Madhya Pradesh where the business can achieve better returns.

**Shift in Product Focus:** Since furniture products contribute less to Flipkart’s overall profitability compared to electronics and clothing, Flipkart should consider reducing its inventory of furniture items and reallocating resources to more profitable categories. Another option could be to streamline its e-commerce focus, concentrating solely on the clothing and electronics sectors, where higher profits are consistently generated.

**Optimize Profit Margins in the West Region:** Flipkart should focus on optimizing profit margins by reassessing its pricing and promotional strategies in the West region. Despite high sales, the profits generated from furniture and clothing are comparable to regions with significantly lower sales, indicating inefficiencies. Flipkart should consider adjusting pricing strategies or reducing costs associated with these products in the West to ensure that high sales translate into higher profits.

**Reflections on Data modeling, Data exploration and Analysis process**

Throughout this assignment, I encountered several new concepts that I hadn't previously covered in other modules, like Data Science Fundamentals. Notably, I delved into data mining and time-series forecasting, which challenged me to identify the best model for profit forecasting, such as Moving Average or Autoregressive. Experimenting with different platforms, I calculated ACF and PACF correlation coefficients, finding that Jupyter Notebook greatly simplified this process with just a single line of code. In week 2, I also learned during the practical about the pareto principle, an important concept in data science, so I implemented it in my dashboard.

Additionally, I revisited Power BI after a year and sharpened my skills in this tool, especially with DAX. Contrary to the notion that this module might be repetitive, I found it quite enriching, particularly with the deeper exploration of K-nearest Neighbour and K-means clustering. The module provided an in-depth understanding of these concepts, such as the iterative process in K-nearest Neighbour that minimizes the distance between observations and centroids. I also leveraged DAX to create new features in Power BI, like customer loyalty ratios, enhancing my proficiency with this Business Intelligence platform.

This assignment allowed me to explore the e-commerce industry's landscape and understand the significance of concepts like cross-selling, which is crucial for business strategy. Lastly, this assignment provided the opportunity to choose my own business question, allowing me to explore real-world scenarios that professionals tackle when addressing critical issues. It broadened my perspective on how data-driven decisions are made in the industry and emphasized the importance of formulating the right questions to uncover insights. This experience challenged me to think beyond the classroom, enhancing my analytical skills and sparking curiosity about the practical applications of data science in solving complex business problems.

# **References**

OpenAI. (2024). GPT-3.5: Generative Pre-trained Transformer.